Weekly Report

2018.0917-2018.0923

1.This Week

Power Grid Deep Learning Paper
1. We keep on training the 2,000 node dataset. When we finished 25 epoch, we find a problem in the data retrieving model which makes the training program reads parts of the data twice or even much more times. We solved this problem and performed modular tests on the code and then restarted the training process. For now, an epoch takes 2 hours (we have 250 epoch in total).
2. I arranged the references, collected KDD paper templates, and tried to compose the paper. Now we have a clear problem but we are somehow lost in what is new for our methods or how can we do better than other methods.
3. We learned the tensor decomposition methods and their mathematical basis. We performed the CP and Tucker decomposition on a small dataset (36 nodes). And I'm trying to figure out how tensor decomposition can help our model.

WaveLines Revision:
1.Revise the waveline paper:
- write a subsection to describe the wavelines presentation and its reason to do that
- remove the data section to introduction
- remove the task section to appendix

Others
1.Prepare for the group meeting presentation. Write the blog for the presented paper.
2.Revise the Dongming's paper.

Working Hour: (except nap and eat time)
8 - 9 hours / week day
2 hours / weekend day
Total Working Hour this week: 46.5 hours.

Paper Reading
1.Latent Variable Time-varying Network Inference (KDD 2018)
This paper proposes a method for graphical modeling of multi-variate time-series. It considers the influence of hidden or unmeasurable factors. The first component of the model represents the connectivity structure of observable variables of the

system. The second component represents the influence of hidden factors (assumed to be few with respect to the observed variables). Experiments show that this model preserves the temporal consistency on both components and provides accurate evolutionary patterns.

$$\underset{\substack{\{\Theta,L\}\\ L_t\succeq 0}}{\text{minimize}} \sum_{t=1}^{T}\Big[-\ell(S_t,\Theta_t-L_t)+\alpha\|\Theta_t\|_{\text{off},1}+\kappa\|L_t\|_*\Big] + \beta\sum_{t=1}^{T-1}\Psi(\Theta_{t+1}-\Theta_t)+\eta\sum_{t=1}^{T-1}\Phi(L_{t+1}-L_t),$$

2.xStream: Outlier Dete‘x’ion in Feature-Evolving Data Streams (KDD 2018)
Feature-Evolving Data Streams: (1) data points may evolve, with feature values changing, as well as (2) feature space may evolve, with newly-emerging features over time.
This paper proposes a density-based ensemble outlier detector. It employs a streaming random projection scheme to embed points in a low-dimensional space on-the-fly (preserving pairwise distances) to handle large and unobserved dimensionality. Then it detects outliers in this projected space via an ensemble of half-space chains. Each half-space chain of the ensemble computes the outlier scores for each point via randomized partitions of the projected input space.

3.Butterfly Counting in Bipartite Networks (KDD 2018)
Butterfly: a complete 2 × 2 biclique in a bipartite graph.
This paper proposed a suite of randomized algorithms that can quickly approximate the number of butterflies in a graph. Edge estimation and vertex estimation are separated in in the algorithm. A global sampling step is used to compute a smaller “sparsified” subgraph, which is used to compute an accurate estimate. It can also randomly sample small subgraphs of the entire graph, and analyze them to compute an estimate.

2.Progress

Work	Deadline	Progress
Power grid paper with Deeping learning	12.15	1.Train the model on the 2,000 node dataset.
SQC Paper	-	1.Delayed
WaveLine revision	ASAP	1.Finished a first version after changing the structure.